Calculating SAT Test Prep Course Return on Investment: A Reanalysis of the Systematic Scientific Literature Review

Ryan Carmichael, MBA

Bruin Financial Management

E-mail: ryan.carmichael@bruinfm.com

November 18, 2019

BACKGROUND AND ANALYSIS

Montgomery and Lilly (hereafter, "M&L") performed a valuable service with their systematic literature review for all SAT coaching studies with sample randomization into treatment and control groups (hereafter, "scientific studies") in order to calculate the average point change. It stands as far and away the highest quality evidence previously available regarding the effectiveness of SAT test preparation. In their study, M&L calculated the weighted average point gain to be 56 points (the weighted mean difference between the treatment groups and their respective control groups) (Montgomery and Lilly 2012). However, a weighted average treatment course hours calculation was mysteriously missing from the entire paper putting a serious limitation on the paper's usefulness. If one is going to calculate the average point gain, then why not calculate the corresponding average hour length that produced that point gain? By not performing the average corresponding course length calculation, the M&L study was but one more consolidated review of multiple studies conducted under the bizarre pretense that the amount of preparation is almost inconsequential and that all courses are equal regardless of their hour length. It is a fact of life that the more you study, the more you learn. Students want to know what length of course they should buy in order to help reach their goal score. Calculating the average point gain, without calculating the corresponding average hour length, is essentially meaningless and useless.

M&L's systematic literature review found 10 scientific studies. The 10 studies they found are also reviewed herein. The studies are: Roberts and Oppenheim (1966), Pike and Evans (1973), Alderman and Powers (1980), Hopmeier (1984), Johnson (1984)[1], Laschewer (1986)[2], Zuman (1988), Shaw (1992), Holmes and Keffer (1995), and McClain (1999). However, three of the

studies (Hopmeier, Zuman, and McClain) were missing data such as raw group means and/or standard deviations and thus could not be included in M&L's weighted mean difference based on variance analysis.

In order to include the experiments that are without standard deviation data, not give undue influence to small experiments of low variance, and give more respect to actual real world outcomes, a regular weighted mean difference based on experiment size was used in the present study. In each of the present study's calculations, n is an experiment's total size (the particular treatment group's size plus the applicable control group's size) and thus its proportional weight within the overall analysis.

All seven studies M&L included in their calculation were complete with both known point change and known hour length and are therefore eligible to be included in the present study's calculations as well, where applicable. Two of the three studies excluded from M&L's calculation likewise could not be included in the present study's calculations either (in Zuman, some students received additional special coaching sessions, therefore making its actual hour length unknown; McClain included no raw group means). Hopmeier, which M&L excluded due to lacking standard deviation data, is however complete with known point change and hour length and is therefore able to be included in the present study's weighted mean difference based on experiment size analysis. Therefore, there are a total of eight studies complete with both known point change and known hour length (hereafter, "complete scientific studies"). Each of the studies administered a different coaching treatment(s) than the others. The eight studies comprise a total of 35 scientific experiments.

Table 1: Weighted Mean Difference of All Complete Scientific Studies

Experiment	n	Proportional Weight	Point Change	Hour Length	Weighted Point Change	Weighted Hour Length
Alderman and Powers (A)	50	0.020210	44.57	3.61	0.900768	0.072959
Alderman and Powers (B)	79	0.031932	-1.04	10.24	-0.033209	0.326985
Alderman and Powers (C)	39	0.015764	-24.26	6.94	-0.382433	0.109402
Alderman and Powers (D)	91	0.036783	-11.62	6.58	-0.427413	0.242029
Alderman and Powers (E)	99	0.040016	14.92	5.74	0.597041	0.229693
Alderman and Powers (F)	72	0.029103	-7.37	3.84	-0.214487	0.111754
Alderman and Powers (G)	94	0.037995	28.54	8.43	1.084382	0.320299
Alderman and Powers (H)	35	0.014147	40.49	44.72	0.572817	0.632660
Holmes and Keffer (1)	34	0.013743	29.83	7.9	0.409951	0.108569
Holmes and Keffer (2)	36	0.014551	48.55	8	0.706467	0.116411
Hopmeier (1 Math)	63	0.025465	37	4	0.942199	0.101859
Hopmeier (1 Verbal)	49	0.019806	36	3	0.713015	0.059418
Hopmeier (2 Math)	62	0.025061	37	4	0.927243	0.100243
Hopmeier (2 Verbal)	51	0.020614	57	3	1.175020	0.061843
Johnson (SF)	35	0.014147	178	30	2.518189	0.424414
Laschewer (1)	27	0.010914	33.51	8.9	0.365711	0.097130
Laschewer (2)	29	0.011722	119.01	8.9	1.395024	0.104325
Pike and Evans (QC)	335	0.135408	90.5	21	12.254446	2.843573
Pike and Evans (DS)	253	0.102264	52.6	21	5.379062	2.147534
Pike and Evans (RM)	244	0.098626	58.6	21	5.779466	2.071140
Roberts and Oppenheim (A)	80	0.032336	7.8	7.5	0.252223	0.242522
Roberts and Oppenheim (B)	32	0.012935	22.7	7.5	0.293614	0.097009
Roberts and Oppenheim (C)	27	0.010914	36.1	7.5	0.393977	0.081851
Roberts and Oppenheim (D)	27	0.010914	15.3	7.5	0.166977	0.081851
Roberts and Oppenheim (E)	32	0.012935	3.8	7.5	0.049151	0.097009
Roberts and Oppenheim (F)	67	0.027082	50.8	7.5	1.375748	0.203112
Roberts and Oppenheim (G)	63	0.025465	46.4	7.5	1.181568	0.190986
Roberts and Oppenheim (H)	31	0.012530	-13.5	7.5	-0.169159	0.093977
Roberts and Oppenheim (I)	32	0.012935	0.4	7.5	0.005174	0.097009
Roberts and Oppenheim (J)	59	0.023848	-1.2	7.5	-0.028618	0.178860
Roberts and Oppenheim (K)	30	0.012126	21.7	7.5	0.263137	0.090946
Roberts and Oppenheim (L)	30	0.012126	17.9	7.5	0.217057	0.090946
Roberts and Oppenheim (M)	32	0.012935	100.3	7.5	1.297332	0.097009
Roberts and Oppenheim (N)	33	0.013339	-1.8	7.5	-0.024010	0.100040
Shaw	122	0.049313	18.52	8	0.913274	0.394503
Total:	2474	1			40.850707	12.419871

The average expected gain for all complete scientific studies found from the weighted mean difference analysis is 40.85 points per 12.42 hours which reduces to 3.29 points per hour.

$(x_{\text{hours}}/12.42) \cdot 40.85 = y_{\text{expected point increase}}$

or

$x_{\text{hours}} \cdot 3.29 = y_{\text{expected point increase}}$

Higher quality prep courses will likely yield higher average gains per hour and lower quality prep courses will likely yield lower average gains per hour.

Each of the complete scientific studies had attrition except for Shaw. Attrition in randomized controlled experiments occurs at random so long as there is not a non-random event that causes it such as a researcher- imposed surprise event. An example of a researcher-imposed surprise event that might cause non-random attrition in a control group would be if the control group was never going to receive the treatment and the sample was not informed of that until after randomization into groups. None of the complete scientific studies disclose anything revealing their attrition was anything but random. Most of the attrition simply appears to be a result of the passage of time and whatever naturally accompanies that. Nevertheless, when Group A has a substantially higher attrition rate than group B, group A's average score will likely be somewhat higher than it would have been had it had the same attrition rate as Group B because Group A had more of its fat trimmed and was likely left with a higher percentage of diligent high achievers. Even in such a scenario, a lot of randomization remains. Also, there are many other reasons that might cause someone to drop out of an SAT prep study such as a busy personal schedule, other more compelling

opportunities arising during wait time, College Board and ETS's statements that coaching does not work, or a bright academic record leading one to believe they do not actually need the prep course to beat the test and/or to get into college. However, on average in prep studies, some students having a lack of academic diligence as time passes would logically seem to be the most frequent contributing factor in random attrition and some would argue almost ever-present to some degree regardless of other contributing factors.

Table 2: Attrition Summaries

<table>
<tr><th>Study</th><th>Summary</th></tr>
<tr><td>Alderman and Powers</td><td>The original size of each specific group was unreported. An aggregate total number of treatment drop outs (32) was discernable and an aggregate total number of control drop outs (48) was discernable, however this obviously does not allow one to make any logical inferences on a per experiment basis.</td></tr>
<tr><td>Holmes and Keffer</td><td>In each experiment, the rates of attrition were too similar between the treatment group and its control group to confidently make logical inferences about mean difference point changes being likely somewhat inflated or likely somewhat deflated. Perhaps with larger group sizes these small differences in attrition rates would allow for such inference, but not with the group sizes present in Holmes and Keffer.
<table>
<tr><th>Group</th><th>Original Size</th><th>Final Size</th></tr>
<tr><td>Treatment 1</td><td>28</td><td>15</td></tr>
<tr><td>Control 1</td><td>30</td><td>19</td></tr>
<tr><td>Treatment 2</td><td>29</td><td>19</td></tr>
<tr><td>Control 2</td><td>28</td><td>17</td></tr>
</table></td></tr>
<tr><td>Hopmeier</td><td>There was no attrition in the math experiments. In each of the verbal experiments, the treatment group had a substantially higher rate of attrition than the control group, thus the mean difference point changes are probably somewhat inflated in the verbal experiments.
<table>
<tr><th>Group</th><th>Original Size</th><th>Final Size</th></tr>
<tr><td>Treatment 1 Verbal</td><td>31</td><td>20</td></tr>
<tr><td>Treatment 2 Verbal</td><td>30</td><td>22</td></tr>
<tr><td>Control</td><td>32</td><td>29</td></tr>
</table></td></tr>
</table>

Table 2: Attrition Summaries (continued)

Study	Summary
Johnson	The control group had a substantially higher rate of attrition than the treatment group, thus the mean difference point change is probably somewhat deflated.

Group	Original Size	Final Size
Treatment	39	23
Control	29	12

Laschewer	The original size of each specific group was unreported. The study began with 80 total participants and the author may have implied that each group began with an equal number, but it is ambiguous. "Eighty students from the original 108 volunteers were randomly selected to participate as subjects in the CAI SAT study. These subjects were randomly assigned to one of the four experimental conditions." (Laschewer 1986:89) In each experiment, the treatment group and its control group ended up being about the same size after attrition.
Pike and Evans	The original size of each specific group was unreported. It is reported that about 8.3% dropped out of the groups in total. This could be a meaningful issue if for example the control group lost 20% of its participants and each of the three treatment groups only lost 3% of their respective participants.
Roberts and Oppenheim	The original size of each specific group was unreported. Although it is stated that the totals in verbal treatment groups, verbal control groups, math treatment groups, and math control groups each started with 150 participants, the actual total in verbal control groups (111) and the actual total in math control groups (122) were each much smaller and the actual total in verbal treatment groups (154) and the actual total in math treatment groups (188) were larger.
Shaw	No attrition.

In a country without a mandated national curriculum, College Board's solution was to design a test to be non-aligned to high school curriculum. In order to make this idea work, they had to also claim the test was uncoachable or else they would be completely hiding the ball from those who could not afford a test prep course to efficiently learn its disparate body of knowledge, not aligned to anything but itself. But if it were proven the SAT was coachable, then its sellability would have been devastated. As such, the results of any study College Board performed or funded could make or

break them, thus creating a conflict of interest of the highest magnitude. This dynamic makes it necessary to also run the weighted mean difference analysis on just the independent complete scientific studies.

Table 3: Weighted Mean Difference of All Independent Complete Scientific Studies

Experiment	n	Proportional Weight	Point Change	Hour Length	Weighted Point Change	Weighted Hour Length
Holmes and Keffer (1)	34	0.066929	29.83	7.9	1.996496	0.528740
Holmes and Keffer (2)	36	0.070866	48.55	8	3.440551	0.566929
Hopmeier (1 Math)	63	0.124016	37	4	4.588583	0.496063
Hopmeier (1 Verbal)	49	0.096457	36	3	3.472441	0.289370
Hopmeier (2 Math)	62	0.122047	37	4	4.515748	0.488189
Hopmeier (2 Verbal)	51	0.100394	57	3	5.722441	0.301181
Johnson (SF)	35	0.068898	178	30	12.263780	2.066929
Laschewer (1)	27	0.053150	33.51	8.9	1.781043	0.473031
Laschewer (2)	29	0.057087	119.01	8.9	6.793878	0.508071
Shaw	122	0.240157	18.52	8	4.447717	1.921260
	508	1			49.022677	7.639764

The average expected gain for all independent complete scientific studies found from the weighted mean difference analysis is 49.02 points per 7.64 hours which reduces to 6.42 points per hour.

$$(x_{\text{hours}}/7.64) \cdot 49.02 = y_{\text{expected point increase}}$$

or

$$x_{\text{hours}} \cdot 6.42 = y_{\text{expected point increase}}$$

Higher quality prep courses will likely yield higher average gains per hour and lower quality prep courses will likely yield lower average gains per hour.

The weighted mean difference analysis was also run on just the College Board/ETS[3] complete scientific studies to compare its points/hours ratio to that of the independent complete scientific studies.

Table 4: Weighted Mean Difference of All College Board/ETS Complete Scientific Studies

Experiment	n	Proportional Weight	Point Change	Hour Length	Weighted Point Change	Weighted Hour Length
Alderman and Powers (A)	50	0.025432	44.57	3.61	1.133520	0.091811
Alderman and Powers (B)	79	0.040183	-1.04	10.24	-0.041790	0.411475
Alderman and Powers (C)	39	0.019837	-24.26	6.94	-0.481251	0.137670
Alderman and Powers (D)	91	0.046287	-11.62	6.58	-0.537854	0.304568
Alderman and Powers (E)	99	0.050356	14.92	5.74	0.751312	0.289044
Alderman and Powers (F)	72	0.036623	-7.37	3.84	-0.269908	0.140631
Alderman and Powers (G)	94	0.047813	28.54	8.43	1.364578	0.403062
Alderman and Powers (H)	35	0.017803	40.49	44.72	0.720829	0.796134
Pike and Evans (QC)	335	0.170397	90.5	21	15.420905	3.578332
Pike and Evans (DS)	253	0.128688	52.6	21	6.768973	2.702442
Pike and Evans (RM)	244	0.124110	58.6	21	7.272838	2.606307
Roberts and Oppenheim (A)	80	0.040692	7.8	7.5	0.317396	0.305188
Roberts and Oppenheim (B)	32	0.016277	22.7	7.5	0.369481	0.122075
Roberts and Oppenheim (C)	27	0.013733	36.1	7.5	0.495778	0.103001
Roberts and Oppenheim (D)	27	0.013733	15.3	7.5	0.210122	0.103001
Roberts and Oppenheim (E)	32	0.016277	3.8	7.5	0.061851	0.122075
Roberts and Oppenheim (F)	67	0.034079	50.8	7.5	1.731231	0.255595
Roberts and Oppenheim (G)	63	0.032045	46.4	7.5	1.486877	0.240336
Roberts and Oppenheim (H)	31	0.015768	-13.5	7.5	-0.212869	0.118260
Roberts and Oppenheim (I)	32	0.016277	0.4	7.5	0.006511	0.122075
Roberts and Oppenheim (J)	59	0.030010	-1.2	7.5	-0.036012	0.225076
Roberts and Oppenheim (K)	30	0.015259	21.7	7.5	0.331129	0.114446
Roberts and Oppenheim (L)	30	0.015259	17.9	7.5	0.273143	0.114446
Roberts and Oppenheim (M)	32	0.016277	100.3	7.5	1.632553	0.122075
Roberts and Oppenheim (N)	33	0.016785	-1.8	7.5	-0.030214	0.125890
	1966	1			38.739130	13.655015

The average expected gain for all College Board/ETS complete scientific studies found from the weighted mean difference analysis is 38.74 points per 13.66 hours which reduces to 2.84 points per hour.

The independent complete scientific studies point change/hours ratio is 126% over that of the College Board/ETS complete scientific studies point change/hours ratio. Sufficed to say, the whole of the College Board/ETS complete scientific studies is not externally validated by the whole of the independent complete scientific studies. However, that does not automatically mean the College Board/ETS studies were fraudulent. Another possibility is that College Board/ETS just happened to pick mostly bad courses on which to base their studies. For instance, the Alderman and Powers study was based on high school teachers doing verbal classes at high schools, rather than on commercially developed courses, and had a weighted point change to hours ratio of just 1.03 points per hour. Nevertheless, the 126% difference combined with the conflict of interest and the fact that the original group sizes were unreported in all three College Board/ETS scientific studies looks really bad.

The fact that the vast majority of scientific research in the area has been limited to very short courses is extremely peculiar, especially when you consider that the oldest and most prominent prep course, Kaplan, began as a 64 hour course in 1946 (a 3.5 month course) (Kaplan and Farris 2001:30-33). Nevertheless, we now know that a 64 hour course would yield an average gain of 410.88 points per the weighted mean difference analysis on all independent complete scientific studies and even an average gain of 210.56 points per the weighted mean difference analysis on all complete scientific studies which even included the College Board/ETS studies.

It should go without saying that as hours get very high, the SAT's 1600 point ceiling will impart a ceiling effect on scores, and as some test takers hit or get close to the test's ceiling, their impact on the points/hours ratio will be reduced. The amount of hours at which the test's ceiling begins to impart a significant ceiling effect will depend on the sample's baseline SAT skills/SAT knowledge.

Finally, a regression analysis weighted for experiment size was performed on each data set in order to gauge the strength of the hours variable at predicting point change outcomes across studies with substantively different treatments.

Table 5: Weighted Regression Analyses

Data Set	*R*	R^2	*F* Significance	Regression Model
All Complete Scientific Studies	0.6039	0.3647	0.000123	y = 2.7x + 7.36
Independent Complete Scientific Studies	0.7837	0.6142	0.007303	y = 5.04x + 10.5
College Board/ETS Complete Scientific Studies	0.6561	0.4305	0.000369	y = 2.81x + .42

The independent complete scientific studies yielded the largest correlation found by the regression analyses (78.37%, a very high correlation for a study in the social sciences on substantively different treatments and randomized controlled experiments). The fact that the mean results of all of these *separate independent* studies correlate together so highly is further evidence that a composite of the independent complete scientific studies is the closest depiction of reality. Despite the independent complete scientific studies administering substantively different treatments, the hour length was able to predict 61.42% of the variance amongst mean scores across the studies. As such, the regression model outputted for the independent complete scientific studies is the most

recommendable and practical equation for savvy consumers out of all of the equations found by the present study.

$5.04x_{\text{hours}} + 10.5 = y_{\text{expected point increase}}$

If someone finds a course they are interested in purchasing, they would find its expected return on investment by simply plugging its number of hours into the independent complete scientific studies regression equation to get the point increase they can expect from the price of the course. Higher quality prep courses will likely yield higher average gains per hour and lower quality prep courses will likely yield lower average gains per hour. Equipped with this knowledge, parents can better avoid egregiously overspending or egregiously underspending on their child's SAT education in effort to achieve their goal score.

HISTORICAL CONTEXT

The results of the present study have sweeping implications for how society understands the US's past, present, and future.

The SAT was created in 1926. Despite occasional tweaks, the SAT has remained essentially the same: a test of math and/or English designed to be not aligned to high school math and English (Kirst 2001; Baird 2012:19-20; Gewertz 2016). Yet the general public's assumption has always

been that it was aligned to high school curriculum and College Board, the owner of the SAT, did much to foster that assumption.

To be clear, the test owner, College Board, is not a college board, but rather a private corporation (University of the State of New York 1957). Although, they have no official capacity, it is difficult to believe their name was chosen without intent to confuse people into believing they have some official governmental authority (while their corporate charter was issued by the state of New York's education department, they are in no way managed by the state of New York) (Dudley 2017). Further, said pseudo-authority, College Board, spread lies about the coachability of the SAT for decades, at least much of the time knowingly so, as shown by the complete scientific studies and College Board's own admissions of staying abreast of the research on the topic.

In 1965 through 1968, College Board published booklets titled "Effects of Coaching on Scholastic Aptitude Test Scores" which they included in pamphlets for counselors and admissions officers titled "College Board Score Reports: A Guide for Counselors and Admissions Officers." Said booklets defined coaching as:

> "a variety of methods used in attempting to increase in a relatively short time students' mastery of the particular skills, concepts, and reasoning abilities tested by the SAT." (College Entrance Examination Board 1965a:4)

And stated that studies have shown that:

> "increases in scores on the SAT that may result from coaching are negligible." (College Entrance Examination Board 1965a:8)

The 1965-1968 guides for counselors and admissions officers went on to say:

> "The evidence collected leads us to conclude that intensive drill for the SAT, either on its verbal or its mathematical part, is at best likely to yield insignificant increases in scores. The magnitudes of the increases which have been found vary slightly from study to study, but they are always small and appear to be independent of the particular method of coaching used and of the level of ability of the students being coached." (College Entrance Examination Board 1965b: 52)

The student guide for 1979-80, titled “Taking the SAT” stated:

> "The verbal and mathematical abilities measured by the SAT develop over years of study and practice. Drilling or last-minute cramming probably will not do much to prepare you for the test." (Educational Testing Service 1979a:3)

In 1979-81, the pamphlet for high School guidance counselors and college admissions officers, titled “ATP Guide for High Schools and Colleges 1979-81” stated:

> "Over the past 25 years, the College Board has conducted many studies on the effect of special preparation programs on SAT score results and has supported the independent investigation of the topic by others. These, studies consistently seem to demonstrate that "coaching," in the sense of intensive drill on sample test questions, does not lead to any significant improvement in students' scores." (Educational Testing Service 1979b:13)

In 1998-2001, College Board's websites stated:

> "Unfortunately, by the time most students begin to worry about admission tests, it's too late to do much about the results. Preparation should begin well before the letters S-A-T or A-C-T are even mentioned." (College Board 1998)

And further went on to state:

> "Some students simply have modest abilities in the areas being tested. Their test scores probably won't improve if they take a special preparation course. In fact, scores might even go down." (College Board 1998)

Perhaps the most troubling and most telling about overall motive is the fact that College Board was willing to broadcast this message of SAT uncoachability at least as late as 2001 despite all of the independent evidence that had come out to the contrary in the 1980s and 1990s.

Additionally, this message of SAT uncoachability was embraced and recited by teachers wanting to avoid accountability as part of their "don't teach to the test" movement. Nevertheless, an occasional teacher or school district would defy the mandates of College Board and the "don't teach to the test" movement by aligning part of their curriculum to the SAT.

Perversely, the US's most famous colleges are the ones that emphasize this exam the most, a dynamic fostered by US "news" companies using this test in college ranking schemes they created in order to make money, create news rather than report it, and perpetuate a social hierarchy based on parental wealth (in collaboration with the colleges themselves who fill out ranking surveys included in the ranking methodology).

And the social hierarchy based on parental wealth has been successfully perpetuated indeed. Five Ivy League universities accept as many applicants from the top 1% of parental incomes as they accept from the bottom 60% of parental incomes combined (Aisch, Buchanan, Cox, and Quealy 2017). Meanwhile, the test prep industry generated $24.57 billion globally in 2016 and is expected to grow to $32.13 billion by 2021 (Chang 2017).

Correlation to grades has always been the test makers' justification that a test measures intelligence (AKA aptitude) (Blum 1978:71). Following that logic, the level of correlation to grades would also be the standard for how *well* a test measures intelligence if that is indeed what one is seeking to measure. A test's ability to predict college grades, especially college freshman grades, has also been called its ability to measure college readiness, college aptitude, and aptitude (AKA intelligence). However, in a massive study of a diverse 77,893 student population, when the SAT

was pit against tests aligned to high school curriculum known as the SAT II (dubbed mere "achievement tests"), the SAT II was found to be an over three times greater measure of aptitude. And after controlling for family income and parental education level, the SAT II was found to be a 12 times greater measure of aptitude than the SAT.[4] It was thus revealed that the SAT is largely a measure of family income and parental education level. High school grades were also found to be a much greater measure of aptitude than the SAT (Geiser and Studley 2001). The present study disentangles the largely confounding variables of family income and parental education level by unequivocally proving that success on the SAT has been highly influenced by the ability to afford a prep course. That is to say, that of the two largely confounding variables of family income and parental education level, only family income has thus far been proven to be able to substantially influence SAT outcomes completely in and of itself as paid commercial coaching courses are traditionally the way SAT coaching courses have been generally available to the public, especially actual course-length courses developed by SAT experts. We know the purpose of the SAT was never to measure intelligence as both achievement tests and high school grades are each far better measures of intelligence. We know the purpose of the SAT was never to measure traditionally acquired knowledge or else College Board would have made effort to align the SAT to high school curriculum. And no one in their right mind would ever argue that the purpose of the SAT was to measure parental education level for the sake of measuring parental education level or as a proxy for inherited intelligence when the SAT has been so conclusively proven to be terrible at measuring intelligence compared to achievement tests. Therefore, the purpose of giving heavy weight to the SAT, or even using the SAT at all, in college admissions can only have been to give an unfair advantage to wealthy students who can afford the prep courses required to learn its disparate body of knowledge not aligned to anything but itself and to hide the ball from poor people because in

order to more easily score in a high percentile you need the masses to score in a low percentile. And that way you keep from ascending to power those who understand what it means to be poor and the daily obstacles it presents, so that the interests of the poor are less likely to be represented in or by the government. It has also been used to screen out those who cannot afford your college's tuition and would require a need based scholarship. It has also been used to screen out the poor from eligibility for the National Merit Scholarship (PSAT). Further, as theorized by Leon Kamin (1974) in "The Science and Politics of IQ" and confirmed by Herrnstein and Murray's (1994) manifesto "The Bell Curve," one of the goals of saying a particular test is essentially uncoachable has always been to portray the poor as unteachable in order to defund public education. And as a proxy for intelligence, Herrnstein and Murray often used, you guessed it, the SAT.

The claim that the SAT is uncoachable is a lie that creates and perpetuates an illusion of meritocracy by pretending that those who score higher are just inherently more intelligent. It is a fascist dictate that sets the rules for interpretation of SAT test results regardless of reality and what their interpretation should be.

Make no mistake, this lie about uncoachable tests predates College Board's adoption of it, but it was adapted for a college entrance exam by Carl Brigham, the SAT's creator and director for the first 10 years of the SAT's use.

In 1923, while individually advocating for the introduction of an SAT type test (a test unaligned to high school curriculum, thus in actuality requiring a prep course to efficiently learn its body of

knowledge), Carl Brigham, in a twisted moment of inspiration came up with the following Orwellian double speak narrative:

> "I, personally, cannot overcome the feeling that our present scheme for limitation of enrollment favors the man with high pecuniary endowment and penalizes many a man with a high native intellectual endowment whose parents cannot afford to send him to a first rate preparatory school or pay his tutoring bills." (Princeton Alumni Weekly 1923)

This audaciously backwards narrative that the SAT was unsusceptible to prep and that it gave an advantage to the poor became the public facing narrative of College Board for much of its existence since hiring Carl Brigham to create the SAT.

Brigham would eventually admit the narrative was a lie in a 1938 newspaper article that did not reference College Board or the SAT:

> "now it is generally conceded that all tests are susceptible to training and to varying degrees of environmental opportunity." (Macdonald 1938)

Unfortunately, for many decades, College Board would continue spreading the lie that coaching cannot significantly raise SAT scores, only finally effectively admitting it was a lie in 2017 (Gewertz 2017).

That lie determined who got what government funding, who got what private investment, who got what jobs, and even who got what bond rating and no doubt greatly contributed to the US's income inequality crisis. In so doing, it also determined what the beneficiaries' children got, what their children's children got, and so on, ad infinitum, and in an ever-expanding sense due to the time-value of money. And it could happen again anywhere.

ENDNOTES

[1] Johnson included three experiments but only the Johnson San Francisco (SF) experiment had sample randomization into treatment and control groups and is therefore the only Johnson experiment included in any of the present study's calculations.

[2] Laschewer reported two sets of totals for each experiment. The present study uses the totals that yield the smaller point gain calculable from Laschewer tables 18 and 21 rather than the totals that yield the larger point gain calculable from Laschewer table 24.

[3] ETS was created in 1947 by members of College Board and two other test owning companies to take over their test administration functions. ETS has been the administrator of the SAT since ETS's creation. ETS and College Board have seemed inseparable for much of the time since ETS's creation with College Board at times funding the studies and publishing the works done by ETS. All of the College Board/ETS complete scientific studies were funded by College Board and performed by ETS.

[4] The SAT II's standardized regression coefficient to college freshman grades was .23 when controlling for high school grades and the SAT. After adding family income and parental education level to the controls, the SAT II's standardized regression coefficient ticked up to .24. The SAT's standardized regression coefficient to college freshman grades was .07 when controlling for high school grades and the SAT II. After adding family income and parental education level to the controls, the SAT's standardized regression coefficient free fell to .02.

REFERENCES

Aisch, Gregor, Larry Buchanan, Amanda Cox, and Kevin Quealy. 2017, January 18. "Some Colleges Have More Students From the Top 1 Percent Than the Bottom 60. Find Yours." *New York Times*. Retrieved September 3, 2019 (https://www.nytimes.com/interactive/2017/01/18/upshot/some-colleges-have-more-students-from-the-top-1-percent-than-the-bottom-60.html)

Alderman, Donald L. and Donald E. Powers. 1980. "The Effects of Special Preparation on SAT-Verbal Scores." *American Educational Research Journal* 17(2): 239–253.

Baird, Katherine. 2012. *Trapped in Mediocrity: Why Our Schools Aren't World-Class and What We Can Do About It*. Lanham, Maryland: Rowman & Littlefield.

Blum, Jeffrey. 1978. *Pseudoscience and Mental Ability: The Origins and Fallacies of the IQ Controversy*. New York, New York: Monthly Review Press.

Chang, Richard. 2017. "Top 3 Trends Affecting U.S. Test Preparation Market Through 2021." *The Journal*, July 6, Retrieved September 3, 2019 (https://thejournal.com/articles/2017/07/06/top-3-trends-affecting-u.s.-test-preparation-market-through-2021.aspx)

College Entrance Examination Board. 1965a. *Effects of Coaching on Scholastic Aptitude Test Scores.* College Entrance Examination Board.

College Entrance Examination Board. 1965b. *College Board Score Reports: A Guide for Counselors and Admissions Officers.* College Entrance Examination Board.

College Board. 1998. "Planning for College: College Admission Testing." Retrieved September 11, 2019 (https://web.archive.org/web/19980423102305/http://collegeboard.com/features/parentgd/html/testing.html)

Dudley, Renee. 2017. "Amid Problems, Specialists Wonder Who's Overseeing Group That Owns SAT." *Reuters*, February 14, Retrieved July 7, 2019 https://www.reuters.com

Educational Testing Service. 1979a. *Taking the SAT.* College Entrance Examination Board.

Educational Testing Service. 1979b. *ATP Guide for High Schools and Colleges 1979-81.* College Entrance Examination Board.

Geiser, Saul and Roger Studley. 2001. *UC and the SAT: Predictive Validity and Differential Impact of the SAT I ad SAT II at the University of California*. Oakland, California: University of California, Office of the President.

Gewertz, Catherine. 2016. "Should SAT, ACT Do Double Duty?" *Education Week*, 35(15): 16-17.

Gewertz, Catherine. 2017. “College Board Reports Score Gains From Free SAT Practice.” *Education Week*, May 8, Retrieved August 20, 2019 (blogs.edweek.org/edweek/high_school_and_beyond/2017/05/college_board_reports_score_gains_from_free_sat_practice.html)

Herrnstein, Richard and Charles Murray. 1994. *The Bell Curve: Intelligence and Class Structure in American Life*. New York, New York: Free Press.

Holmes, Thomas and Ronald Keffer. 1995. “A Computerized Method to Teach Latin and Greek Root Words: Effect on Verbal SAT Scores.” *Journal of Educational Research* 89(1): 47–50.

Hopmeier, George H. 1984. “The Effectiveness of Computerized Coaching for Scholastic Aptitude Test in Individual and Group Modes.” PhD dissertation, College of Education, Florida State University.

Johnson, Sylvia T. 1984. *Preparing Black Students for the SAT – Does It Make a Difference? (An Evaluation Report of the NAACP Test Preparation Project)*. New York: National Association for the Advancement for Colored People.

Kamin, Leon. 1974. *The Science and Politics of I.Q.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Kaplan, Stanley and Anne Farris. 2001. *Test Pilot: How I Broke Testing Barriers for Millions of Students and Caused a Sonic Boom in the Business of Education.* New York, New York: Simon and Schuster.

Kirst, Michael W. 2001. *Overcoming the High School Senior Slump: New Education Policies.* Perspectives in Public Policy: Connecting Higher Education and the Public Schools.

Laschewer, Arnold D. 1986. "The Effect of Computer Assisted Instruction as a Coaching Technique for the Scholastic Aptitude Test Preparation of High School Juniors." PhD dissertation, Hofstra University.

Macdonald, W.A. 1938. "Brigham Adds Fire to 'War of I.Q.'s'." *New York Times*, December 4, D10.

McClain, T. Benjamin. 1999. "The Impact of Computer-Assisted Coaching on the Elevation of Twelfth-Grade Students' SAT Scores." Doctoral dissertation, Morgan State University.

Montgomery, Paul and Jane Lily. 2012. "Systematic reviews of the effects of preparatory courses on university entrance examinations in high school-age students." *International Journal of Social Welfare* 21: 3-12.

Pike, Lewis W. and Franklin R. Evans. 1973. "The Effects of Instruction for Three Mathematics Item Formats." *Journal of Educational Measurement* 10(4): 257–272.

Princeton Alumni Weekly. 1923. Princeton University Press, November 28, 185-187.

Roberts, S.O. and Don B. Oppenheim. 1966. *The Effect of Special Instruction Upon Test Performance of High School Students in Tennessee.* Princeton, New Jersey: Educational Testing Service.

Shaw, E. 1992. "The Effects of Short-Term Coaching on the Scholastic Aptitude Test" Doctoral dissertation, University of La Verne.

University of the State of New York. 1957. The University of the State of New York Education Department Absolute Charter of College Entrance Examination Board.

Zuman, John P. 1988. *The Effectiveness of Special Preparation for the SAT: An Evaluation of a Commercial Coaching School*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.